Проблемы согласованности моделей в микросервисной архитектуре







Валерий Разномазов

- Руководитель группы внедрения и разработки корпоративных бизнес-приложений
- Кандидат физико-математических наук

Технологический центр Accenture в России
+7-904-342-35-86
valerii.raznomazov@accenture.com

Случай из жизни:
три банка объединяются в один



«Семантическое безумие» в ИТ-системах

«номер счета физического лица»

«20-значный номер счета физика»

«номер отделения, в котором открыт счет физического лица»

«номер счета физика»

«12-значный счет физического лица»

«счет физического лица»



Это всё одно и то же?

Содержание



«Семантическое безумие» в ИТ-системах

Часто при аналитике ИТ-систем мы говорим разными словами об одном и том же, и одинаковыми словами — о разном





Проблема > Решение > Цель

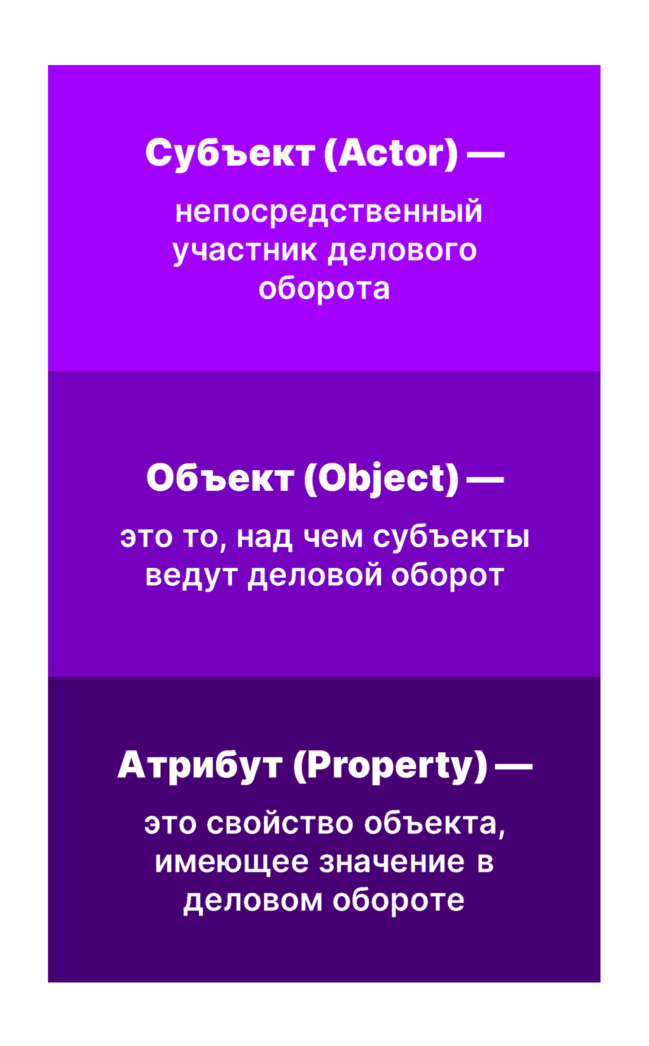
Проблема	«Семантическое безумие»
Решение	Определить субъекты и объекты предметного поля и сопоставить им классы и таблицы баз данных в информационной системе
Цель	Получить объекты с набором атрибутов, которые бы однозначно характеризовали его в любом необходимом бизнес-процессе

«Видов столько, сколько различных форм в начале мира создал Всемогущий»

— Карл Линей (1707 - 1778)



Бизнес-субъекты и бизнес-объекты



Деловой оборот банка

Бизнес-объект

IT-система

Передача	Обработка	Хранение
Сообщение в информационной системе	Java-class	Таблица в базе данных

JAVA

Схемы XSD и JSONschema

```
<?xml version="1.0"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema">

<xs:element name=«Contractor">
  <xs:complexType>
    <xs:sequence>
      <xs:element name="FullName" type="xs:string"/>
      <xs:element name="AccountNumber" type="xs:string"/>
      <xs:element name="address" type="xs:string"/>
      <xs:element name="phone" type="xs:string"/>
    </xs:sequence>
  </xs:complexType>
</xs:element>

</xs:schema>
```

```
"employees": {
"description": "Contractor",
"type": "array",
"uniqueItems": true,
"items":
          { "type": "object",
          "properties":
                    { "FullName": {
                              "type": "string",
                     },
                    "AccountNumber": {
                              "type": "string"},
                    "address":{
                              "type":"string"},
                    "phone":{
                              "type":"string"} },
          }
 }
```

Классы языков программирования и таблицы баз данных

```
public class Customer {
private String Fullname;
private Double AccountNumber;
private String Address;
Private String Phone;
}
```

```
CREATE TABLE CUSTOMERS(
  ID   INT                  NOT NULL,
  FullNAME VARCHAR (120)     NOT NULL,
  AccountName  VarChar(20)            NOT NULL,
  ADDRESS  CHAR (25) ,
  Phone   Char (18),
  PRIMARY KEY (ID)
);
```

«Семантическое безумие» —

состояние артефактов не позволяющее однозначно определить сематический смысл передаваемых данных

Дублирование данных в сообщениях и таблицах

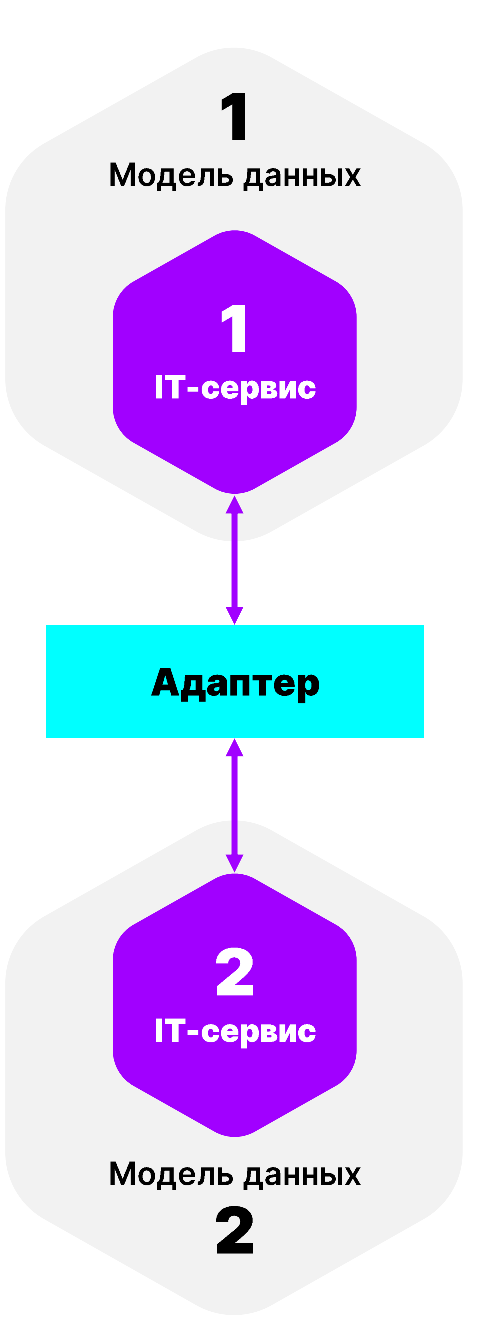
Невозможность привести таблицы баз данных данных к нормальным формам

Затруднения при определении границ микросервисов

Отсутствие единства и согласованности терминологии в документации

Интеграция с внешней системой

Случай №1	Интеграция с внешней системой
Что нужно сделать	Необходимо сделать маппинги, по которым в адаптере будут осуществляться преобразование данных из одной модели в другую
В чём сложность	Однозначно сопоставить одну сущность другой не всегда представляется возможным
Почему	Команды пользуются различной терминологией и различным сленгом, который используют для описания процессов



Семантическое безумие в Confluence

Случай №2	Семантическое безумие в Confluence
Что нужно сделать	Необходимо описать процесс, происходящий на платформе понятным всем работающим командам языком
В чём сложность	Если Wiki страница редактируется членами различных команд, имеющих различный сленг и использующих различную терминологию, есть риск потерять исходный смысл описываемой сущности
Почему	Команды пользуются различной терминологией и различным сленгом при описании процессов

Я называю этот процесс в Camunda «Вилами»

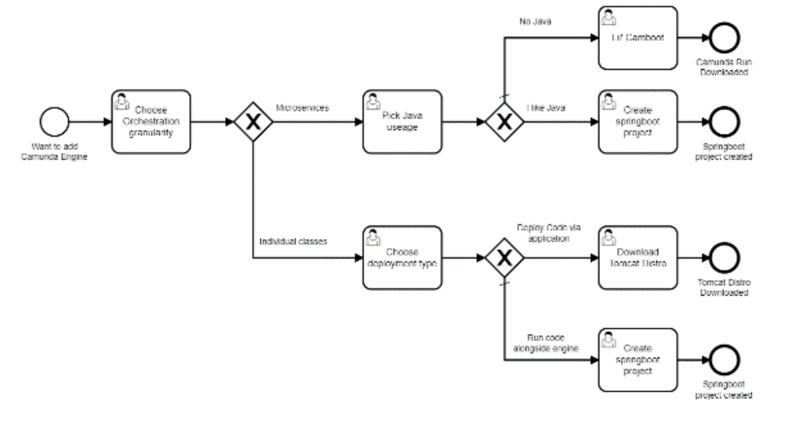
Я называю этот процесс в Comunda «Триггером»

А я говорю, что это «шлюз»

Да это же старый добрый If-else

Больше похоже на switch

…

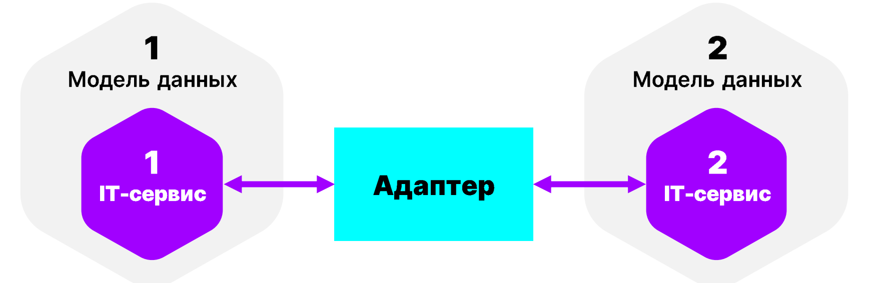




Поглощение одного бизнеса другим

Случай №3	Поглощение одного бизнеса другим
Что нужно сделать	Необходимо интегрировать одну IT-систему в другую
В чём сложность	Однозначно сопоставить одну сущность другой не всегда представляется возможным
Почему	Команды пользуются различной терминологией и различным сленгом при описании процессов





Новая команда заходит на Платформу

Случай №4	Новая команда заходит на Платформу
Что нужно сделать	Необходимо интегрировать новую команду в IT-систему
В чём сложность	Однозначно сопоставить одну сущность другой не всегда представляется возможным
Почему	Команды пользуются различной терминологией и различным сленгом при описании процессов


IT Arch

Модель данных




Analyst

Справочники, словари




PM

Нотации




Developer

Фреймворки и СУБД




Tester

Стратегия тестирования



Глобальные бизнес-объекты и онтологии





Определение объектов и субъектов предметного поля

Декларативный способ

Что это?

Аналитики заранее договариваются о субъектах, объектах и их атрибутах, действуя совместно

Когда?

До начала работ

Метод

Декларативное описание предметного поля на уровне верхнеуровневой архитектуре

Построение онтологии

Что это?

Построение онтологии при помощи онтологического моделера

Когда?

Работы уже начаты

Метод

Построение онтологической модели предметного поля при помощи языка OWL

Кластеризация

Что это?

Применение экспрессных, как правило математических методик для определения глобальных бизнес-сущностей, учавствующих в информационном обмене

Когда?

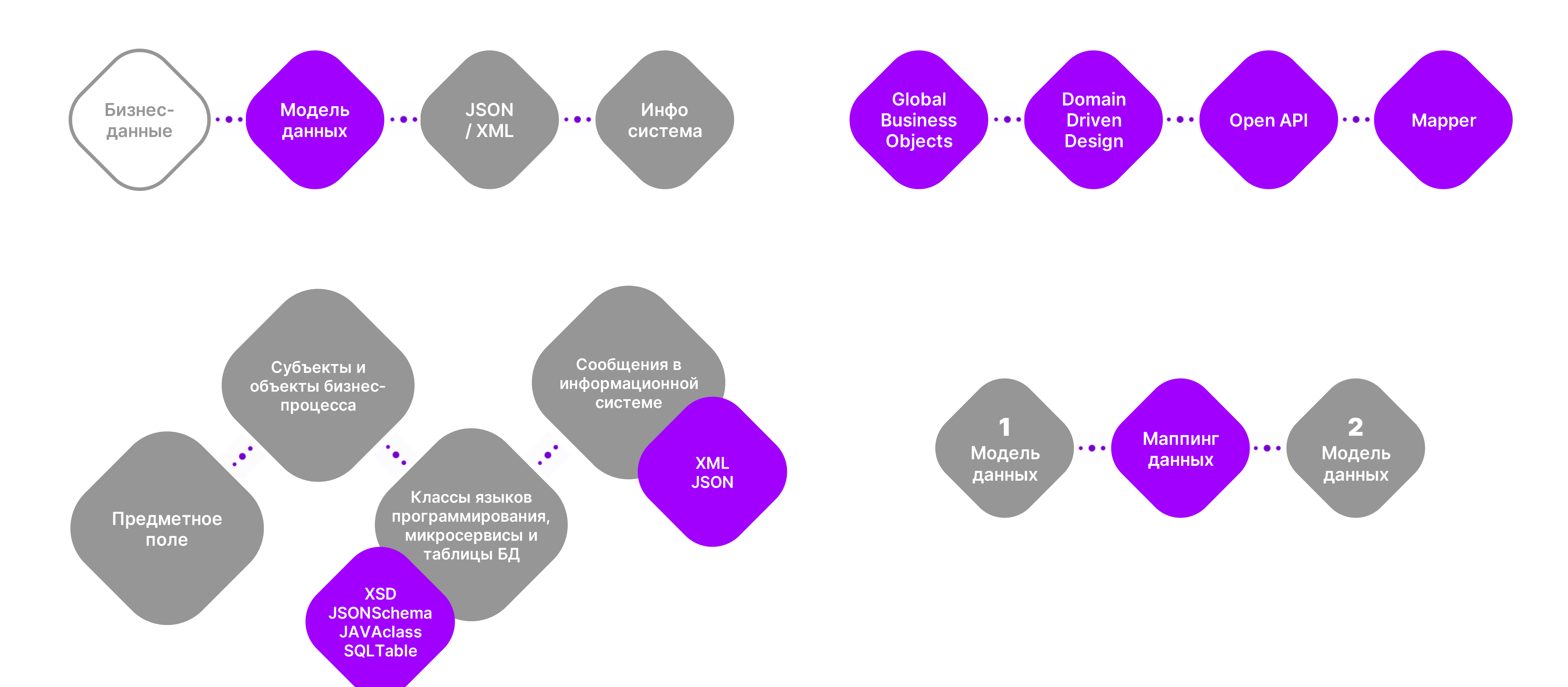
Работы уже начаты

Метод

Разрабатывается алгоритм или группа алгоритмов анализа артефактов и семантического описания объектов и их атрибутного состава



Определение объектов и субъектов предметного поля



Формально верхнеуровневая спецификация

Необходимо интегрироваться с уже работающей системой и суть передаваемых данных известна. Структура данных в этом случае может быть скопирована и схемы применены аналогичные.

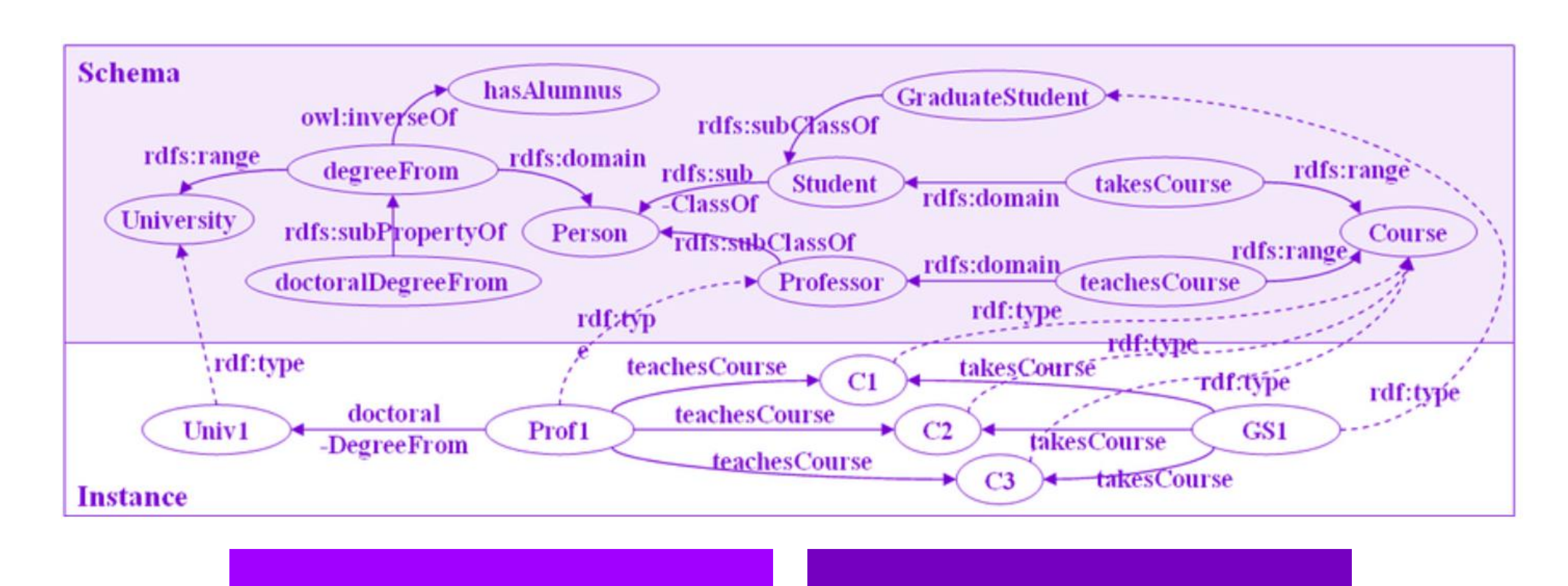
Документация | **HLA** | **Confluence**

Онтология

Онтология

Семантическая разметка	Кластеризация
Анализ семантики языка	Определение субъектов, объектов и связей между ними
Определение семантического ядра	Построение концептуальной схемы взаимодействия
Определение глобальных сущностей на его основе	

Построение онтологической модели при помощи онтологических моделеров и языка OWL



JSONSchema

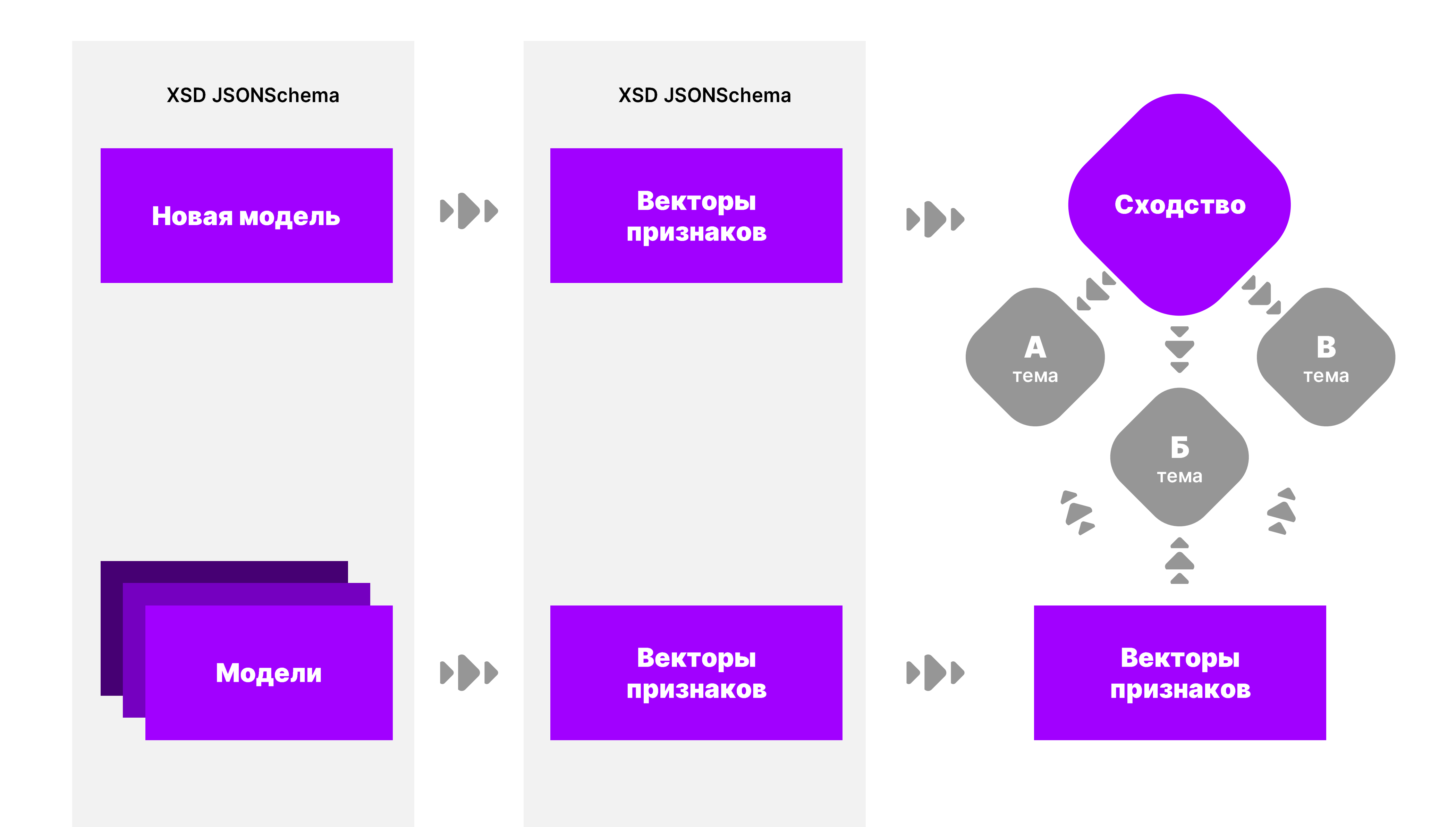
XSD

Методы кластеризации для приведения моделей данных





Конвейер кластеризации



Метрики пространственного сходства

Сходство строк

Дистанция редактирования

- Левенштейна
- Смита-Ватермана
- Аффинное

Выравнивание

- Джаро-Винклера
- Soft TF-IDF
- Монг-Элкан

Фонетика

- Созвучие
- Перевод
- x

Метрическое расстояние

- Евклидово
- Манхетенское
- Минковского

Аналитика текста

- Жаккарда
- TF-IDF
- Косинусово сходство

Сходство отношений

На основе множеств

- Кубика
- Жаккара
- До общих соседей
- Взвешенное расстояние Адара

Группировка

- Средние значения
- Максимальные\ минимальные значения
- Медиана
- Частота (мода)

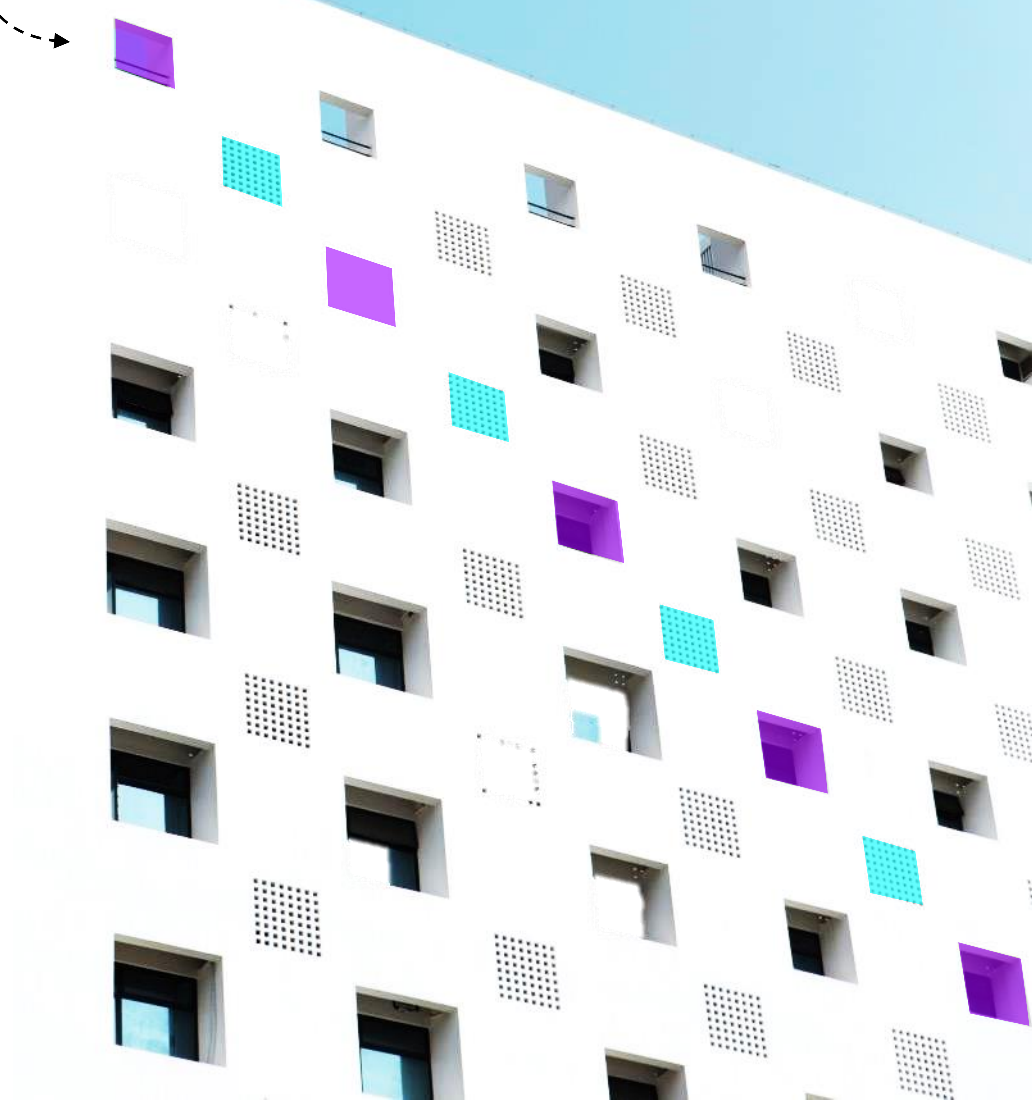
Алгоритм Левенштейна. Поиск редакционных расстояний

«Номер счета физического лица»

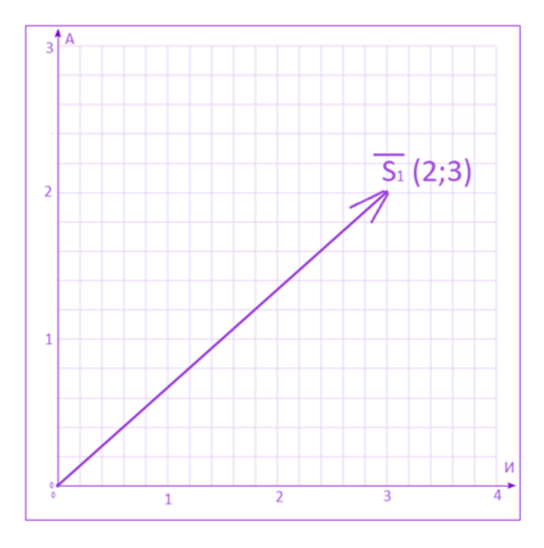
«Номер счета физика»

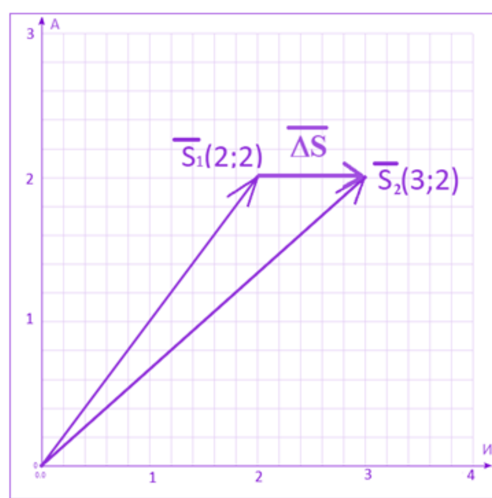
«Номер счета юридического лица»

Редакционное расстояние



«Номер счета физического лица»





$$\boldsymbol{\overline{S_1} = (2; 0; 0; 1; 0; 3; 0; 0; 1; 3; 0; 1; 1; 1; 1; 3; 0; 1; 2; 1; 0; 1; 0; 1; 2; 0; 0; 0; 0; 0; 0; 0; 0)}$$

«Номер счета физического лица»

$$\overline{S_1} = (2; 0; 0; 1; 0; 3; 0; 0; 1; 3; 0; 1; 1; 1; 1; 3; 0; 1; 2; 1; 0; 1; 0; 1; 2; 0; 0; 0; 0; 0; 0; 0; 0)$$

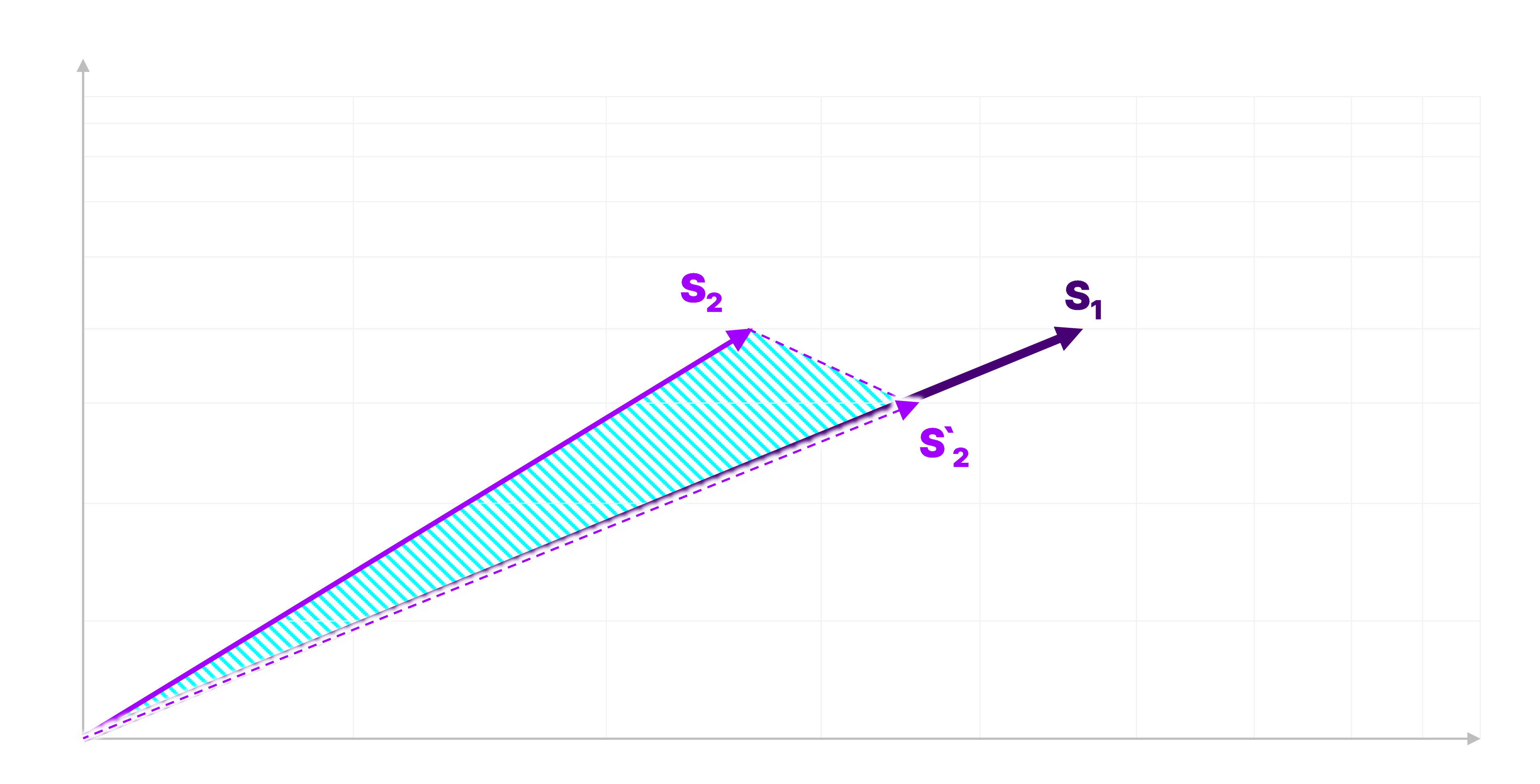
«Номер счета «физика»

$$\overline{S_2} = (2; 0; 0; 0; 0; 2; 0; 0; 1; 2; 0; 1; 0; 1; 1; 1; 0; 1; 1; 1; 0; 1; 0; 0; 1; 0; 0; 0; 0; 0; 0; 0; 0)$$

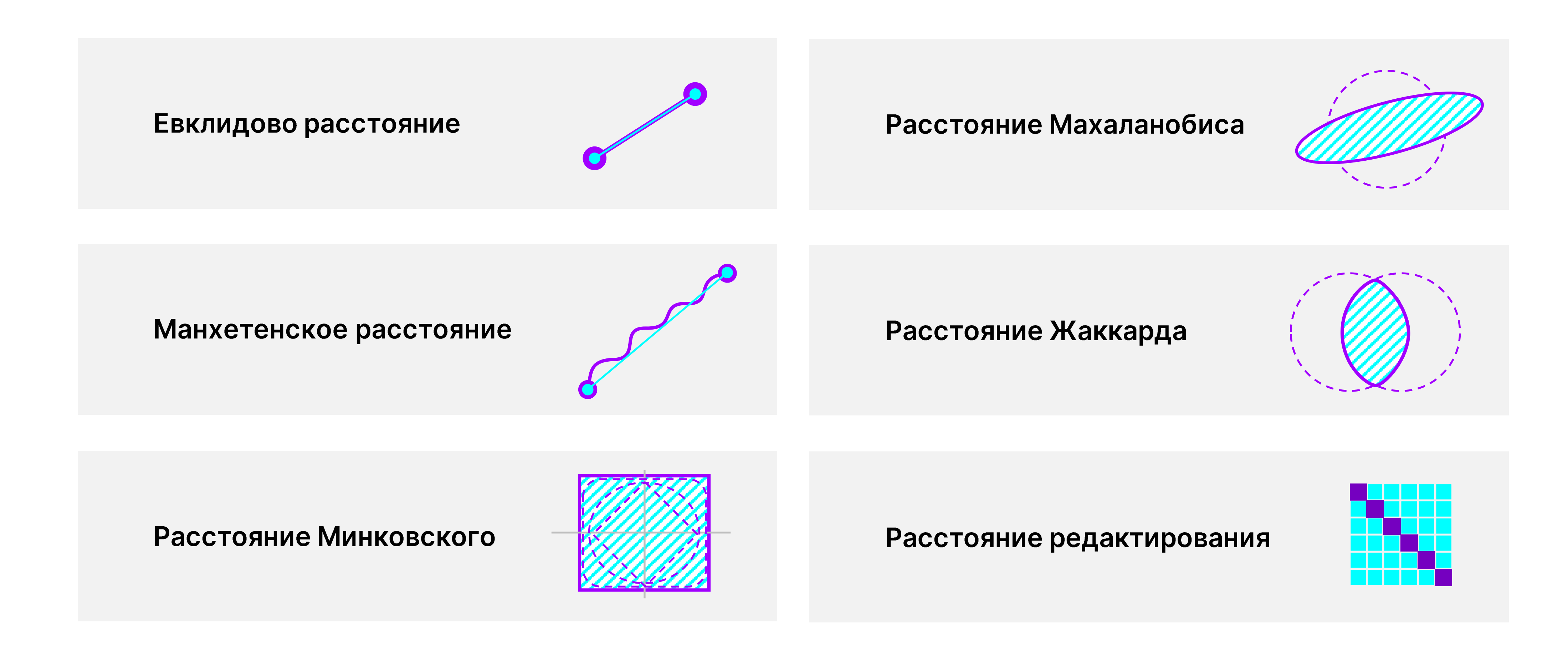
Разница между векторами

$$\overline{\Delta S_{1,2}} = (x_{А,1} - x_{А,2}; x_{Б,1} - x_{Б,2}; \ldots\ldots; x_{Я,1} - x_{Я,2})$$

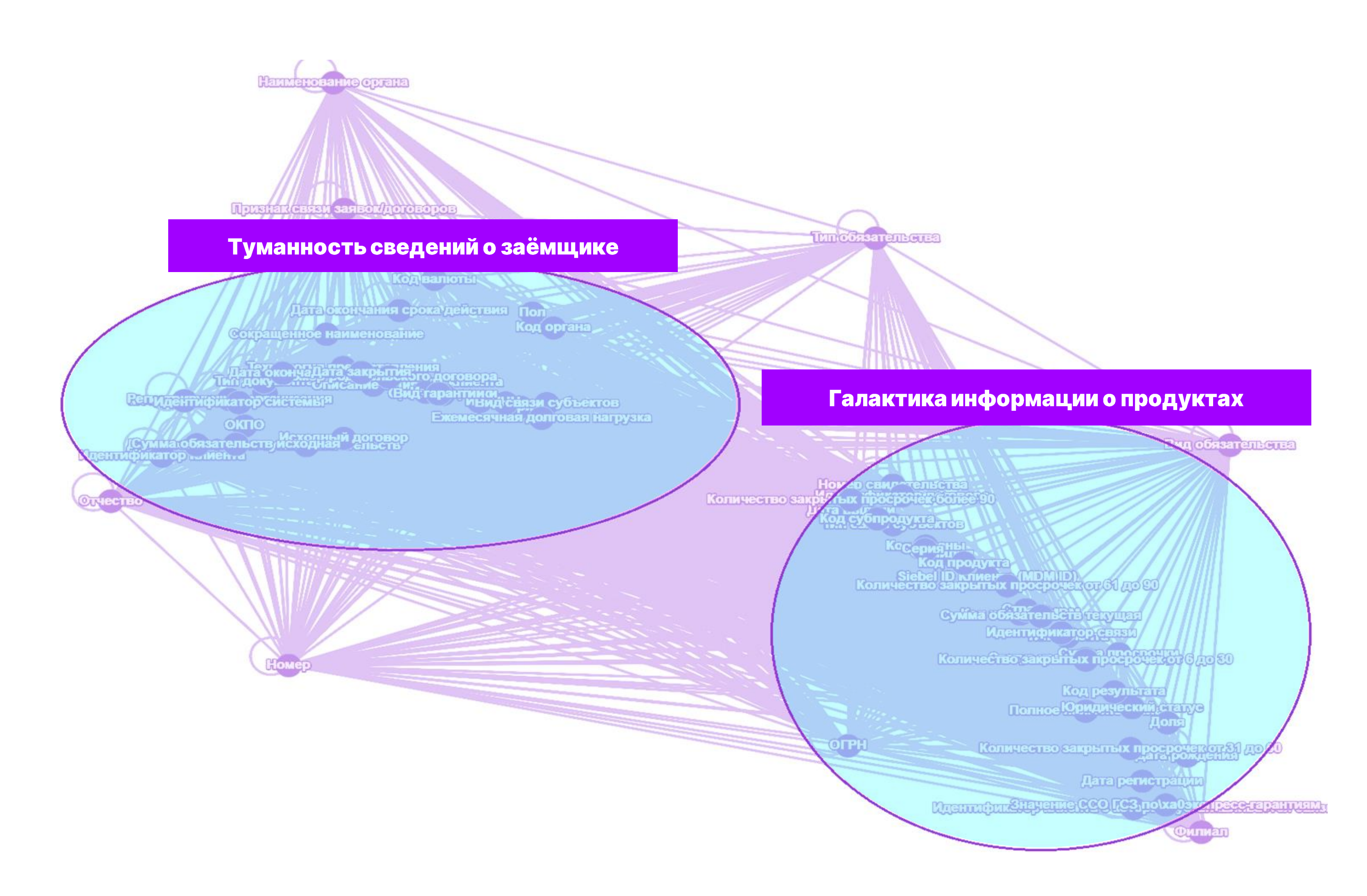
Учет угла между векторами. Проекция одного вектора на другой



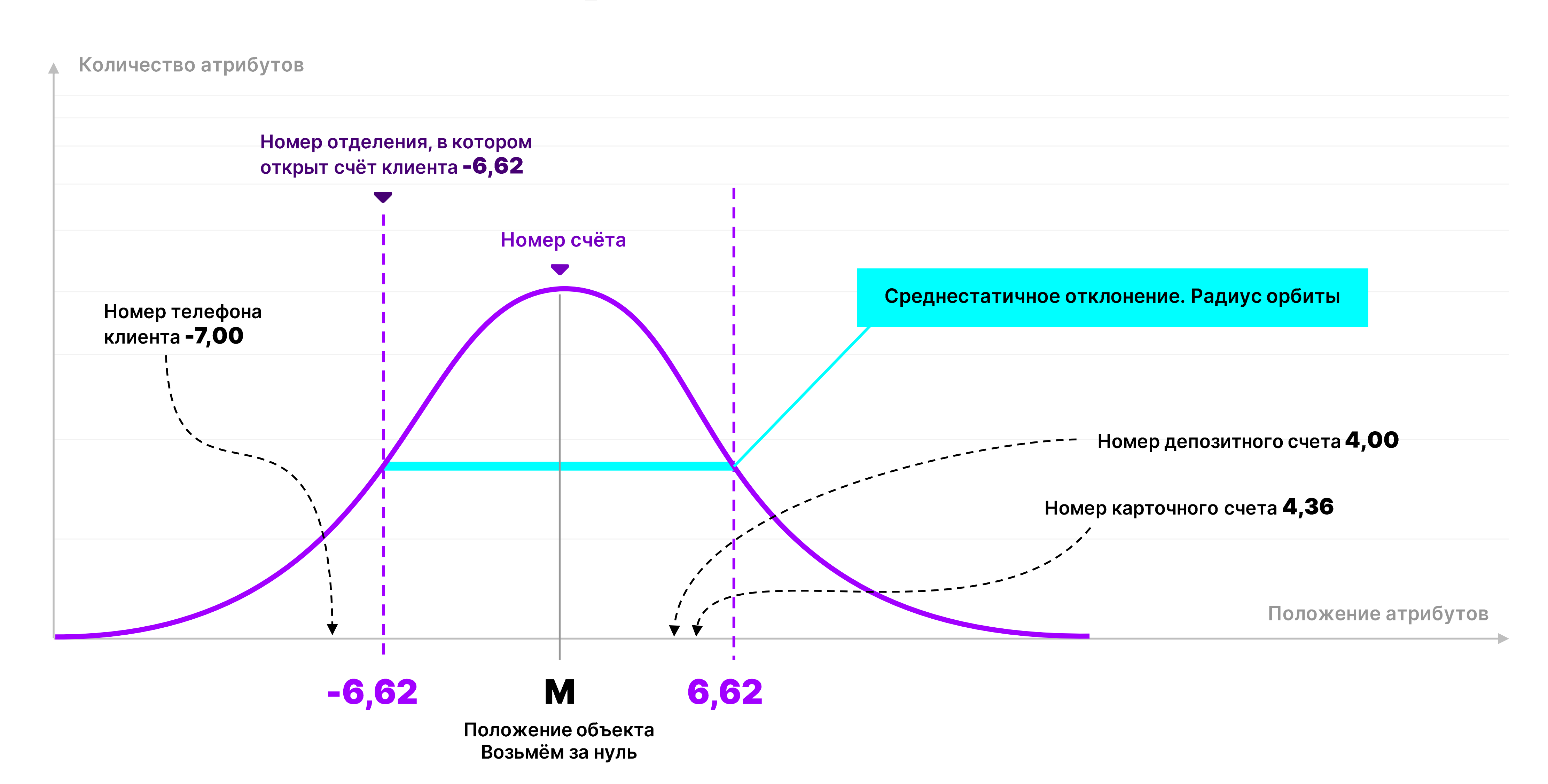
Учет угла между векторами. Проекция одного вектора на другой



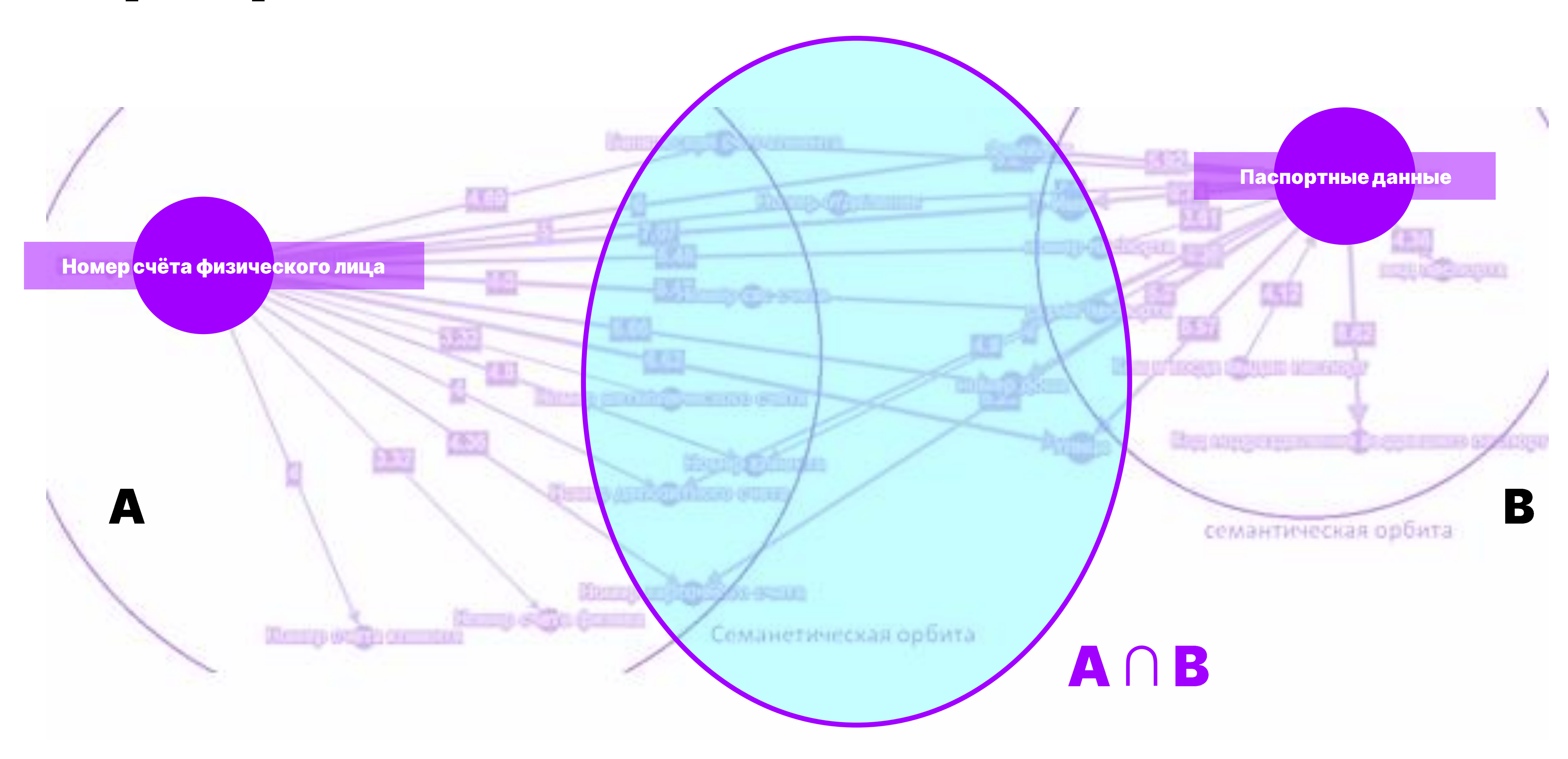
Учет угла между векторами. Проекция одного вектора на другой



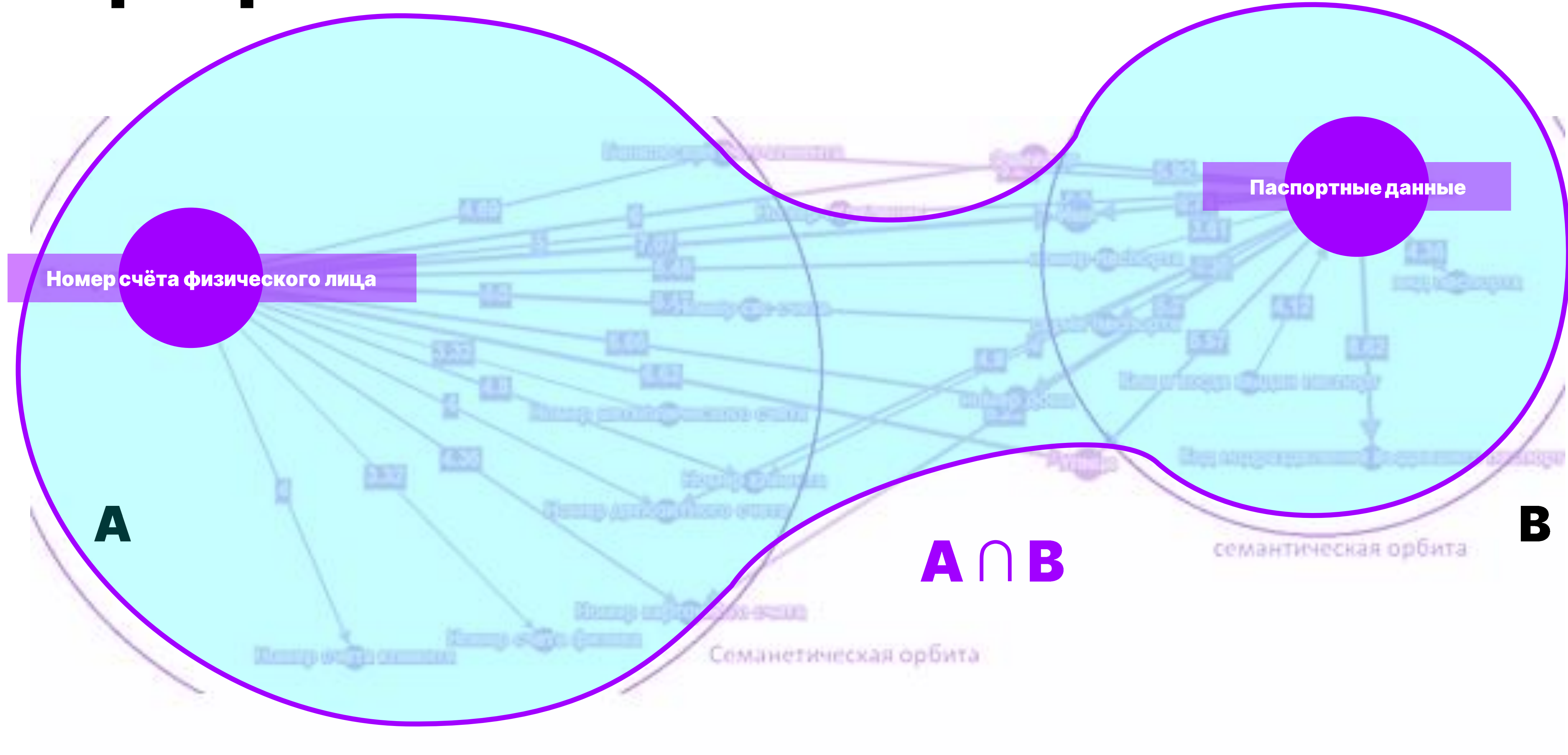
Как определить величину семантической орбиты?



Визуализируем предметное поле через расстояние Минковского

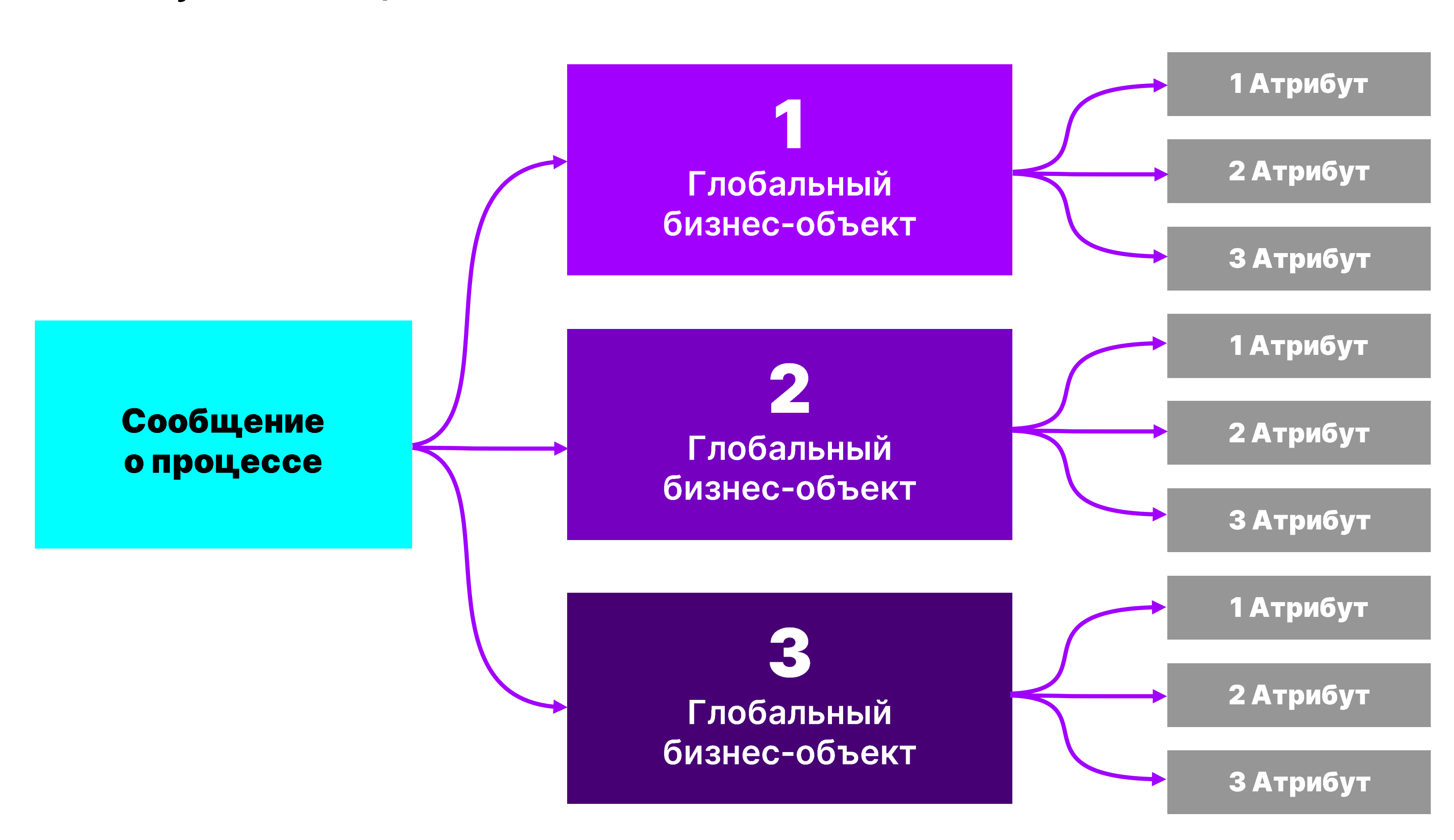


Визуализируем предметное поле через расстояние Минковского



Практическое использование общей схемы бизнес-объектов при работе с гибкими методологиями и микросервисами

Используем в сообщениях глобальные бизнес-объекты

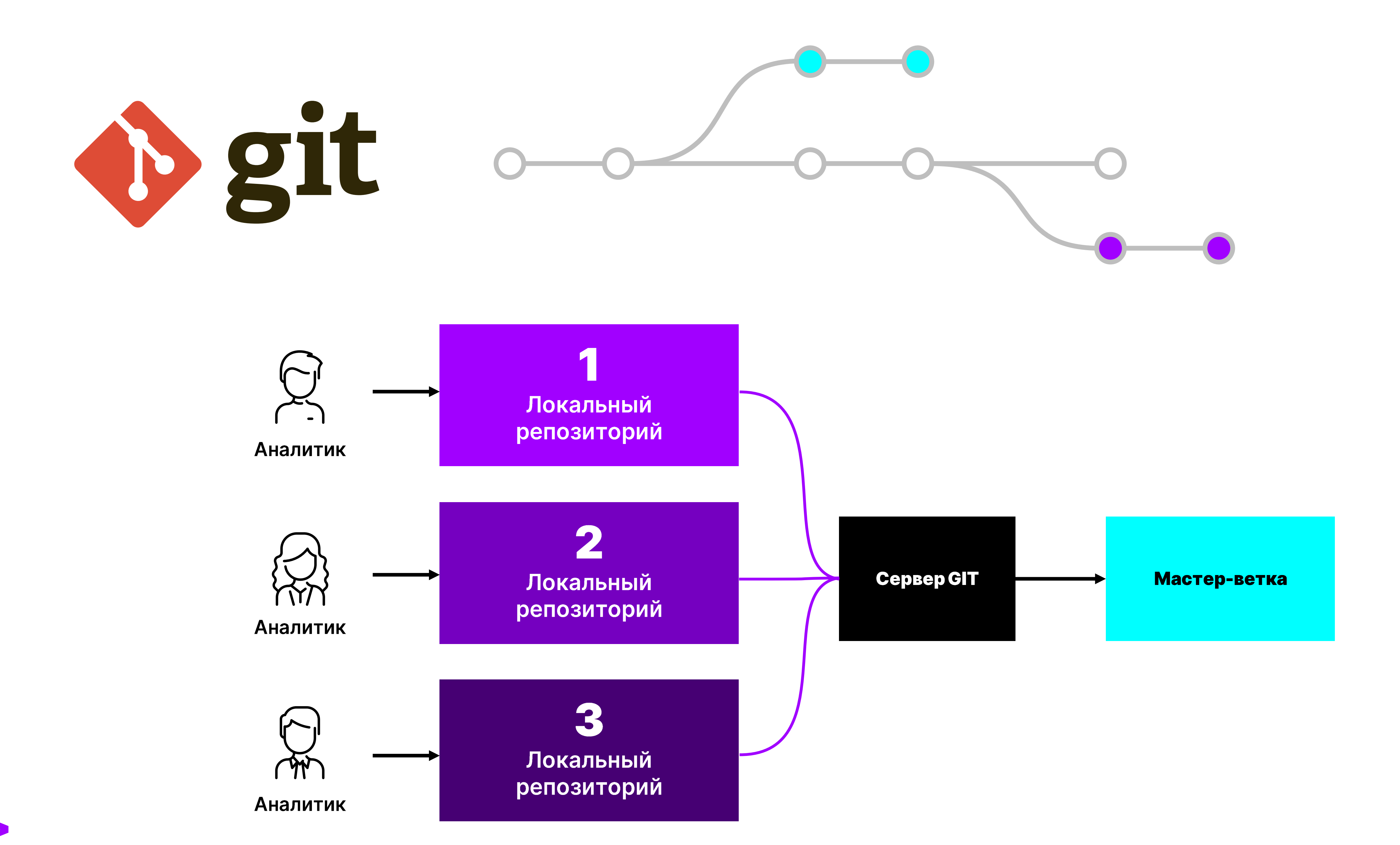


Структурированное для передачи данных набирается из глобальных бизнес-объектов

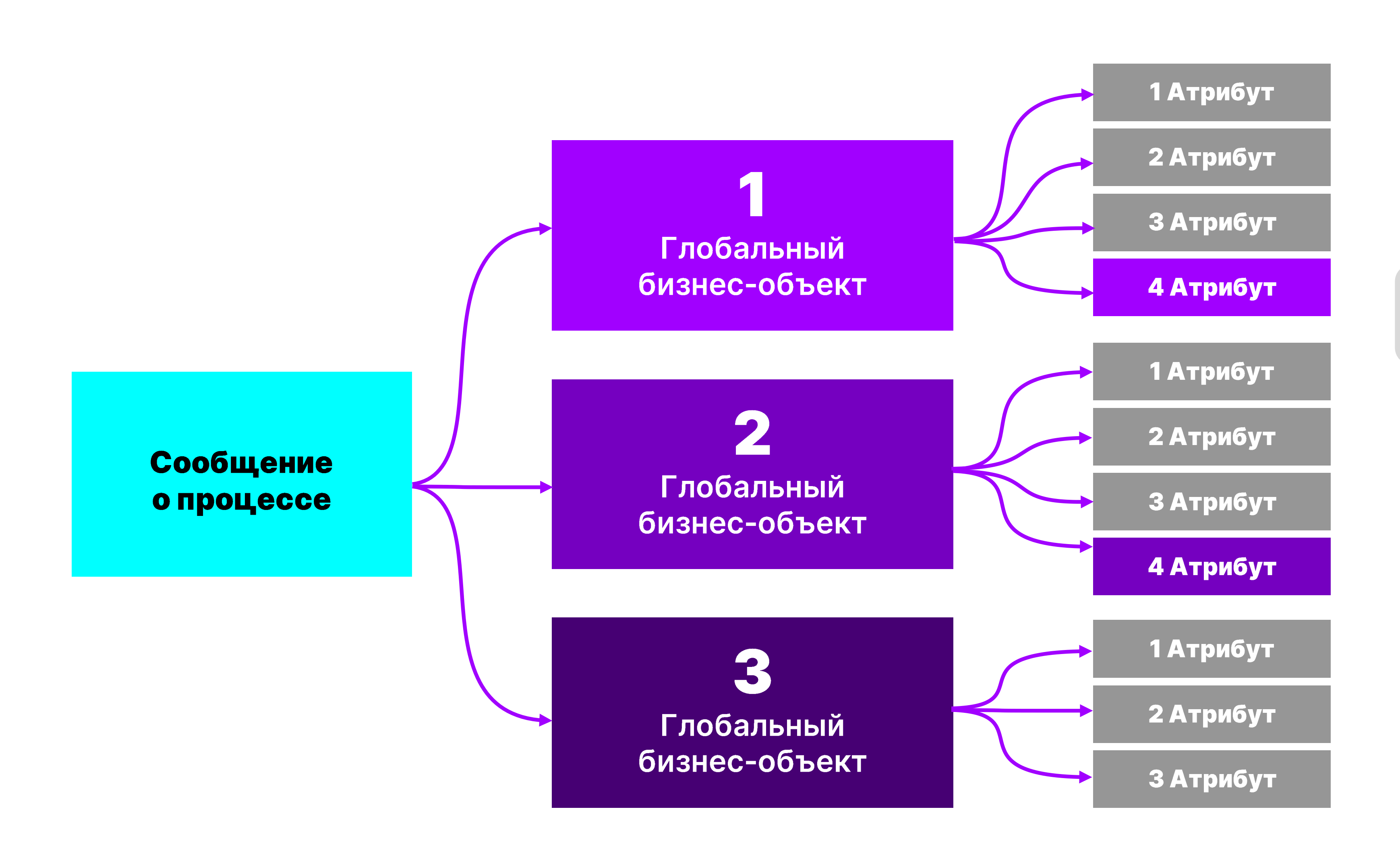
Во избежание ошибок всем атрибутам выставляется обязательность **FALSE**

Административный ресурс




git
Аналитик
1
Локальный репозиторий
Аналитик
2
Локальный репозиторий
Аналитик
3
Локальный репозиторий
Сервер GIT
Мастер-ветка

Меняем модель данных путем добавления к глобальным бизнес-сущностям новых атрибутов



По мере надобности новые атрибуты добавляются в конец старых

Старые атрибуты не удаляются никогда

Что же дальше?



Что же дальше?

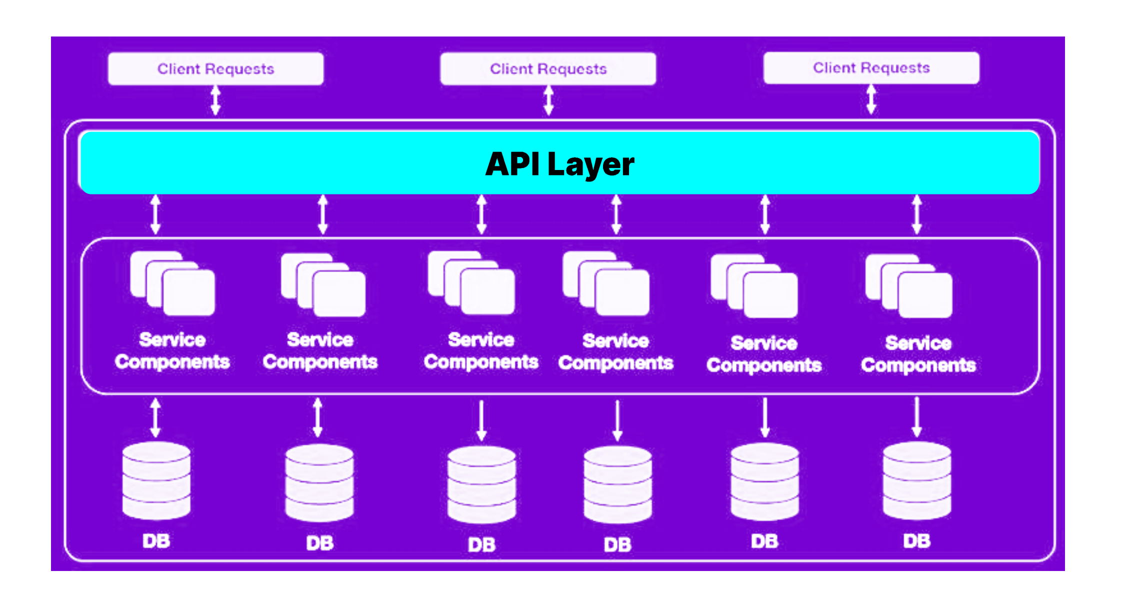
Интернет вещей IoT

Интеграция всех со всеми

Стандартизация и классификация

Границы микросервисов

Случай №5	Семантическое безумие порождает размытые границы микросервисов
Что нужно сделать	Определить объекты и субъекты бизнес-процессов
В чём сложность	Из-за различной терминологии и семантики атомизировать процессы не всегда представляется возможным
Почему	Команды пользуются различной терминологией и различным сленгом при описании процессов



Что же дальше? Краткие выводы.

1

Составлем онтологический словарь предприятия и на его основе делаем API и мапперы

2

Используем статистические и математические методы для кластеризации и оценки схоимости моделей данных

3

Для передачи данных и построение табл используем глобальные бизнес-сущности



Спасибо за внимание _



